



Matematisk Statistik

SF1911 Statistik för bioteknik: Lp 2 2017

Lab 1 för
CBIOT3

Introduktion

Detta är handledningen till Laboration 1, ta med en en utskriven kopia av den till laborationen. Läs handledningen två gånger. Försäkra dig om att du förstår hur de MATLAB-kommandon som finns i den bifogade koden fungerar. Laborationen bedöms som godkänd eller ej godkänd. För att få delta i laborationen skall svar på förberedelseuppgifter kunna redovisas **individuellt**. Arbete i grupp är tillåtet med **högst två** individer per grupp. Godkänd laboration ger 2 poäng till ordinarie tentamenstillfälle.

Syfte och introduktion

Du kan skriva dina kommandon direkt i MATLAB-prompten men det är absolut att föredra att arbeta i editorn. Om den inte är öppen så kan du öppna den och skapa ett nytt dokument genom att skriva `edit lab1.m`. Koden som ges nedan är skriven i celler. En ny cell påbörjas genom att skriva två procenttecken. `Ctrl+Enter` exekverar innehållet i en cell.

Problem 1 - Simulering av slumpstal

Vi börjar med att simulera slumpstal ifrån en normalkurva/fördelning.

Plotta en standardiserad normalkurva d.v.s. klockkurvan med $\mu = 0, \sigma = 1$, med `plot` funktionen och `normpdf`.

Dra sedan 1000 slumpstal ifrån normalkurvan med `normrnd`.

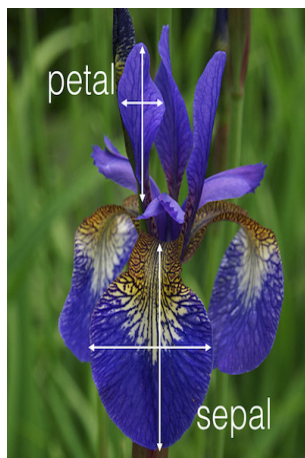
Skapa ett histogram av slumpstalen med funktionen `histogram`. Histogrammet använder dock absoluta tal och inte frekvenser så för att jämföra mot den riktiga fördelningen så är det lättare att använda `ksdensity`. Den jämnar också ut kurvan. Med hjälp av kommandot `hold on` så går det att rita flera gånger i samma figurer.

Hur förhåller sig histogrammet till fördelningen och hur förklaras eventuella skillnader?

Problem 2 - Undersökning av Iris blommor

I denna del av laborationen kommer ni att undersöka ett dataset bestående av tre arter av blommor i Irissläktet. Ni kan ladda in datat i matlab med `load fisheriris`, det ger er två variabler, `meas` och `species`. Varje rad motsvarar en individuell planta och varje kolumn i `meas` motsvarar en variabel. Variabeln `species` innehåller arten som plantan tillhör. De första 50 plantorna är av arten *setosa*, följt av 50 stycken tillhörande arten *versicolor* och de sista 50 tillhör arten *virginica*. Kolumnerna i `meas` variabeln motsvaras av följande variabler:

1. Sepal length
2. Sepal width
3. Petal length
4. Petal width



En blomma med sepal och petal markerade. [Wikimedia commons](#)

Börja med att göra boxplots för varje kombination av variabel och art. Få med alla arter i en figur så att ni kan jämföra mellan dem för den variabeln. Boxplots görs med funktionen `boxplot`. *Subplot* och *figure* kan också vara användbara.

Hur tolkas en boxplot? Vilka slutsatser kan ni dra om arterna och de olika variablerna?

Gör histogram av alla variabler i `meas` med `ksdensity`. Är någon variabel som följer normalkurvan (är normalfördelad)? Det kan vara lättare att se om de följer normalkurvan med `qqplot`.

Plotta alla variablerna `meas` mot varandra med `gplotmatrix`. Vilka variabler ser ut att vara korrelerade? Beräkna korrelationskoefficienten för de som ser ut att vara korrelerade, detta kan göras med hjälp av funktionen `corr`.